

[\[Next\]](#) [\[Back to MUC-6 main page\]](#)

Tokenization Rules

Text elements annotated for the MUC-6 Named Entity and Coreference tasks must consist of one or more complete tokens. Normally, the presence of whitespace surrounding a single character or a group of characters defines an explicit token (a word). This document explains where boundaries of tagged strings are meant to be located when there is NO explicit whitespace between alphanumeric characters and a punctuation mark or other special character.

Named Entity tagging is used in this document as an example of the effects of tokenization. The tokenization rules apply also to the Coreference and Information Extraction tasks.

1 - Punctuation and special characters are normally considered separate tokens.

2 - When a proper name or number contains an internal punctuation mark or other special character, the word containing that character is treated as just one token.

3 - Hyphen at end of line

Tokenization Rules - 14 JUN 95

[\[Next\]](#) [\[Back to MUC-6 main page\]](#)

Generated with [CERN WebMaker](#)



[Homepage](#) | [Advanced Search](#)

Search using:

CUSTOM WEB FILTERS

[Tools](#) | [HotBot Skins](#) | [Pref](#)
Date: **Before December 5 2000** [[Edit this Search](#)]

SPONSORED LINKS (filters not applied)

- [Teragram Entity Extractor](#)

Entity & Event **Extraction** Software For Enterprises, Portals and Media
www.teragram.com

- [Tip: Web Data Extraction](#)

Automate your web tasks with our Award-winning iOpus Internet Macros
www.iopus.com/iim.htm

- [Data Extraction Tool](#)

Extraction and integration of web data with efficient GUI-based tool.
www.kapowtech.com

- [Information Extraction](#)

advanced platform for transforming text into structured DB **information**
www.celi.it

- [Advanced Data Extraction](#)

Highly cost-effective **extraction**. Any data on the web.
www.poorva.com/aie

[Save on Web Data](#)
 One-time or scripter
 can afford, Fortune
extradata.com/

[Information Extr:](#)
 Towards Scalable, A
 Systems Only \$49.9
 used).
Amazon.com

[VisualText](#)
 Build it. NLP, IE, pa
 crawlers, text data
www.textanalysis.c
 Plac

WEB RESULTS (Showing Results 1 - 10 of 25,823)

1. [MUC-6](#)

MUC-6, the sixth in a series of Message Understanding Conferences, was held in November 1995. ... w
 involved the evaluation of **information extraction** systems applied to a common task, have...
cs.nyu.edu/cs/faculty/grishman/muc6.html - April 24, 1996 - 6 KB

2. [RALI -- Bilingual Information Extraction](#)

... The task of an **information extraction** system is to identify specific **information** from a natural la
 text ... a car accident report, an **information extraction** system will be able to ...
www.iro.umontreal.ca/~kousseim/Extraction/ProjetEI.en.html - June 12, 2000 - 4 KB

3. [Towards Translingual Information Access using Portable Information...](#)

... using Portable **Information Extraction**. Michael White, Claire Cardie, Chung-hye Han, Nari Kim, ..
 the feasibility of combining. portable **information extraction** with MT in ...
www.cogentex.com/papers/Towards-TransIA-using-PIE.pdf - May 7, 2000 - 100 KB

4. [Wrapper Induction for Information Extraction](#)

University of Washington. Computer Science & Engineering. Research on Wrapper Induction for **Infor**
Extraction ... resources typically use hand-coded wrappers, customized procedures for **information** ...
www.cs.washington.edu/homes/weld/wrappers.html - February 16, 1998 - 2 KB

5. [Information Extraction Supported Question Answering](#)

... **Information Extraction** Supported Question Answering ... This paper discusses the use of our **inf**
extraction (IE) system, Textract, in the question- ...
trec.nist.gov/pubs/trec8/papers/cymfony.pdf - February 7, 2000 - 45 KB

6. Information Extraction: New Developments in Astronomical Information...

... this **information** to the proper distributed resource. **Information extraction** is a really complex .. will illustrate the complexity of the **information extraction** process ...

www.eso.org/gen-fac/libraries/lisa3/lestevens.html - September 21, 1998 - 27 KB

7. Designing a Mixed-Initiative Information Extraction System

Designing a Mixed-Initiative **Information Extraction** System. Peter Vanderheyden. and. Robin Cohen of Computer Science. University of Waterloo. Waterloo, Ontario. Canada N2L 3G1. fpbv ander. r cohen: oo:ca. Abstract ... mixed-initiative **information extraction** system. formed, a user formulates a query the ...

www.cs.wright.edu/people/faculty/mcox/mii/papers/vanderheyden.pdf - December 18, 1999 - 102 KB

8. Automatically Constructing a Dictionary for Information Extraction...

... created by AutoSlog will achieve strong performance for **information extraction** from novel text: achieve good performance. on **information extraction** tasks for limited domains ...

www-nlp.cs.umass.edu/ciir-pubs/riloff_aaai93.pdf - October 11, 1999 - 59 KB

9. Information Extraction

Information Extraction. There are a number of text processing applications where the emphasis lies to process large quantities of real text (e.g. ... sometimes also referred to as message understanding, scanning, or **information extraction** ...

www.ltg.ed.ac.uk/ltg_muc/ltg_muc.html - April 16, 1996 - 4 KB

10. 1 Overview of Information Extraction Task

... Next] [Previous] [Top] [Back to MUC-6 main page] **Information Extraction** Task Definition ... The of the **Information Extraction** (IE) task is to provide ...

cs.nyu.edu/cs/faculty/grishman/IEtask15.book_2.html - June 15, 1995 - 7 KB

« [Previous](#) | [Next](#) »

Search for "**information extraction**" using: [Google](#), [Ask Jeeves](#)

Portions powered by  **Inktomi**

[Advertise](#) | [Help](#) | [Text-only Skin](#) | [Submit Site](#) | [HotBot International](#) | [Yellow Pages](#)

© Copyright 2004, Lycos, Inc. All Rights Reserved. | [Privacy Policy](#) | [Terms & Conditions](#) | [HotBot Your Site](#)
